Credit Card Fraud Detection using Machine Learning

1. Project Title

Credit Card Fraud Detection using Machine Learning

---

2. Problem Statement

Credit card fraud is a serious issue in the financial sector where unauthorized transactions lead to financial loss and loss of customer trust. Since fraudulent transactions are very rare compared to genuine ones, traditional rule-based systems struggle to detect fraud effectively. This project aims to solve this problem using machine learning techniques.

---

3. Objective

The main objective of this project is to build a machine learning model that can accurately identify fraudulent credit card transactions while minimizing false alerts for genuine users.

---

4. Dataset Description

The dataset contains historical credit card transactions made by users. Each record represents a transaction with multiple numerical features generated for privacy reasons. The target column indicates whether a transaction is fraudulent (1) or genuine (0).

Key points:

- Highly imbalanced dataset
- Fraud cases are very less compared to normal transactions

- Numerical features only

---

5. Data Preprocessing

The following preprocessing steps were performed:

- Checked for missing values

- Feature scaling for better model performance

- Handled class imbalance using appropriate techniques

- Split the dataset into training and testing sets

---

6. Exploratory Data Analysis (EDA)

Exploratory analysis was performed to understand the data distribution:

- Analyzed fraud vs non-fraud transaction ratio

- Observed transaction patterns

- Identified imbalance in the target variable

EDA helped in selecting suitable evaluation metrics instead of relying only on accuracy.

---

7. Model Selection

Logistic Regression was used as the baseline machine learning model because it is simple, interpretable, and effective for binary classification problems.

---

8. Model Training

The model was trained using the training dataset. Proper hyperparameters were chosen to ensure stable learning and to avoid overfitting.

---

9. Model Evaluation

The model was evaluated using the following metrics:

- Accuracy

- Precision

- Recall

- F1-score

- ROC-AUC

These metrics were selected because accuracy alone is not sufficient for imbalanced datasets like fraud detection.

---

10. Results

The model achieved strong performance across all evaluation metrics, indicating a good balance between detecting fraud and avoiding false alarms. High precision ensured fewer genuine transactions were flagged, while high recall ensured most fraud cases were captured.

---

11. Improvements

Possible improvements to the project include:

- Using advanced models like Random Forest or XGBoost

- Adding behavior-based features

- Optimizing decision thresholds

- Using real-time transaction data

---

12. Future Scope

- Real-time fraud detection system

- Continuous model retraining with new data

- Integration with banking systems

- Explainable AI for better transparency

---

13. Tools & Technologies Used

- Python

- Pandas

- NumPy

- Scikit-learn

- Jupyter Notebook

---

14. Conclusion

This project demonstrates how machine learning can be effectively used to detect credit card fraud. By using appropriate evaluation metrics and handling data imbalance, the model provides reliable fraud detection and has strong potential for real-world deployment.

---

15. References

- Credit Card Fraud Dataset from KAGGLE

- Scikit-learn Documentation

- Machine Learning research articles